

Decision on Intercensal Population Estimates

March 12, 2003

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Decision on Intercensal Population Estimates

The Census Bureau will produce intercensal population estimates based on the official Census 2000 results rather than on a population base adjusted using the revised estimates of census coverage. The Accuracy and Coverage Evaluation (A.C.E.) Revision II methodology represents a dramatic improvement from the previous March 2001 A.C.E. results. However, several technical concerns remain, including uncertainty about the adjustment for correlation bias, errors from synthetic estimation, and inconsistencies between demographic analysis estimates and the A.C.E. Revision II estimates of the coverage of children. Given these technical concerns, the Census Bureau has concluded that the A.C.E. Revision II estimates should not be used to change the base for intercensal population estimates.

INTERCENSAL POPULATION ESTIMATES

The Census Bureau produces annual data on the population size and certain population characteristics (age, race, ethnicity, and sex) of the nation, states, and counties. In addition, Title 13, Section 181 of the U.S. Code requires the Census Bureau to produce biennial estimates of total population for all local units of general purpose government, regardless of their size. Further, the law specifies the use of such estimates by federal agencies when allocating federal benefits to states, counties, and local units of government when those benefits are based on population size.

Among the federal programs that use these intercensal estimates to allocate funds are the Department of Health and Human Services' Medical Assistance Program (Medicaid) and Social Service Block Grant Program, the Department of Housing and Urban Development's Community Development Block Grant Program, and the Department of Labor's Employment and Training Assistance-Dislocated Worker Program. About \$200 billion in federal funds is distributed annually to states and other areas based in some part on intercensal estimates.

These estimates of the geographic distribution of the population are also used for decisions about providing state and local government services, planning utility services, redefining metropolitan areas, and locating retail outlets and manufacturing establishments. Federal time-series statistics that are produced on a per capita basis, such as per capita income, births per capita, and cancer incidence rates, also rely on these estimates for their denominators. Finally, they are used as population controls for the major household surveys and, hence, have a major impact on the accuracy of the country's key indicators such as employment and unemployment, inflation, household income, poverty, and health insurance coverage.

The Census Bureau produces intercensal population estimates for about 3,000 counties, 19,000 incorporated places, and 17,000 functioning minor civil divisions. These entities include a large number of areas with fewer than 10,000 total population (see Table 1). As we develop these estimates, they are shared with the representative of each state for review and comment. This cooperative and collaborative process is essential to developing population estimates that are a reliable and useful indicator of how the population in the United States changes between censuses. Once the intercensal estimates are released, the highest elected official in each area has the right to challenge the estimates for that area through a designated challenge process.

CHOICE OF THE BASE POPULATION FOR INTERCENSAL ESTIMATES

After each census, the intercensal estimates program revises the population base to reflect the results of the most recent census. Intercensal population estimates throughout the following decade result from incorporating estimates of population change based on vital statistics (for births and deaths) and administrative records (for migration) into this base population. Intercensal population estimates developed for 2001 did not include an adjustment to correct for estimated net coverage error in Census 2000, nor did the 2002 national and state estimates released in December 2002.

Since that time, the results from the Accuracy and Coverage Evaluation (A.C.E.) Revision II have become available. The rest of this document provides the rationale for the decision not to use these results to change the base for the intercensal population estimates.

RESULTS OF A.C.E. REVISION II

With the recent work on A.C.E. Revision II, the Census Bureau now has a much better understanding of the influences on census coverage. Moreover, this recent work dramatically improves measures of net coverage compared with the March 2001 A.C.E. estimates (see the *Technical Assessment of A.C.E. Revision II*). The results of A.C.E. Revision II are substantially different from those of March 2001, changing the estimated net coverage of the total household population from a net undercount of 1.18 percent to a net overcount of 0.49 percent.

A.C.E. Revision II estimated a net overcount of 1.13 percent for non-Hispanic Whites, but a net undercount of 1.84 percent for non-Hispanic Blacks. Net coverage estimates for all other race/Hispanic origin groups (Hispanics, Non-Hispanic Asians, Native Hawaiians and Other Pacific Islanders, American Indians and Alaska Natives on Reservations, and American Indians and Alaska Natives off Reservations) were not statistically different from zero (see Table 1 of the *Technical Assessment of A.C.E. Revision II*).

A.C.E. Revision II estimates that about 80 percent of the 19,269 incorporated places had net census overcoverage — with net overcounts of more than 2 percent in 35 percent of places and net overcounts of 0 to 2 percent in 45 percent of places. In contrast, only 2 percent of places had net undercounts of greater than 2 percent and 18 percent of places had net undercounts of 0 to 2 percent. The results vary greatly by size of place, with smaller proportions of larger places showing estimated net overcounts and higher proportions of smaller places having estimated net overcounts of 2 percent or more (see Table 2).

ASSESSING THE RESULTS OF A.C.E. REVISION II

The October 2001 report of the Executive Steering Committee for A.C.E. Policy proposed investigating whether using revised coverage estimates that addressed the major problems of the March 2001 estimates could improve the intercensal population estimates, with particular attention to reducing the differential coverage error in Census 2000. A review of the results and evaluations now available leads us to conclude that, although accuracy might be improved on average by adjusting intercensal estimates, troubling anomalies and unexplained results mean that the Census Bureau cannot be confident of improvements in accuracy at the levels of geography for which estimates are produced.

The *Technical Assessment of A.C.E. Revision II* identifies a number of limitations of the results, many of which are of greater concern for subnational than for national estimates. (Recall that the intercensal estimates program requires population estimates for the nation, states, and counties by age, sex, race, and ethnicity, as well as estimates of total population for small places.) Four aspects of the technical limitations are particularly salient to the decision not to change the base for the intercensal estimates: uncertainty about the adjustment for correlation bias, errors from synthetic estimation and choice of post-strata, inconsistencies with demographic analysis results for children, and the incompleteness of the total error model.

Adjustment For Correlation Bias

One of the long-standing criticisms of the standard dual-system models upon which census coverage estimates have been based is that they do not account for the phenomenon that people who were missed in the census may be more likely to be missed in the post-enumeration survey than those whom the census counted. For example, some potential respondents may deliberately avoid being counted in either the census or survey to avoid contact with government agencies. Others may live in unconventional housing units that are more likely to be missed by both the census and the A.C.E.

For A.C. E. Revision II, correlation bias was corrected to the extent possible using sex ratios obtained from demographic analysis results. Because the demographic analysis results are limited to the two race categories – Black/non-Black – and are available only at the national level by age and sex, several options were available to implement the correlation bias adjustment across the various post-strata.

Using any of these options produces the same results for the total population, but these results differ considerably from those obtained with no correction for correlation bias. The A.C.E. Revision II total net overcount estimate of 0.49 percent would be a net overcount estimate of 1.12 percent without the correlation bias adjustment. More dramatically, for non-Hispanic Blacks, the A.C.E. Revision II estimated net undercount of 1.84 percent would be an estimated net overcount of 0.53 percent without the correlation bias adjustment. These results, and results for other race groups, are shown in Table 3.

Concern was expressed about two aspects of the correlation bias adjustment. One concern was the uncertainty about the appropriate model for allocating among post-strata the correlation bias estimated at the national level for the age-race (Black/non-Black) groups. The middle three columns of Table 3 provide some illustration of how results can vary across three alternative models. (The estimates for non-Hispanic Blacks at the national level are direct, involving no allocation, so they do not vary across the alternative models.) The largest variation in the estimates occurs for Native Hawaiians and Other Pacific Islanders, although the estimates for this group have large standard errors due to a relatively small sample size. Results from column five of Table 3 are discussed below.

The second concern about the correlation bias adjustment is the assumption of no correlation bias for children and adult women. The comparisons between the A.C.E. Revision II estimates and the demographic analysis results discussed in a later section are relevant to the issue of correlation bias for children. The assumption of no correlation bias for adult women was dictated by the decision to use only the sex ratios obtained from demographic analysis to serve as the basis for estimating correlation bias. The totals from demographic analysis were not viewed as being sufficiently reliable, in general, to use to estimate correlation bias for both adult men and adult women. While comparisons of the demographic analysis totals with the A.C.E. Revision II results lend some support to the assumption that the major problem with correlation bias occurs for adult men, particularly for Blacks, they are only rough indications. It is certainly possible that correlation bias exists for adult women, including adult women in some subgroups of the non-Black population, but the demographic analysis results provide no direct indications for such groups.

In particular, concern exists about the possible level of correlation bias for Hispanics. Because a large percentage of the Hispanic population are foreign born, and some are in the United States without appropriate documentation, avoidance of government contact may lead to substantially higher correlation bias for Hispanics—both women and men—than for most other non-Blacks. As an indication of this possibility, column five of Table 3 (“modified two-group model”) shows results obtained by assuming that the correlation bias for Hispanics equals that for Blacks. This model is not strictly within the methodology used by the other models (in that results for Hispanics are not obtained from data on non-Blacks, but rather are borrowed from the results for Blacks), and the large male-female differential implied by the results for Blacks is not the issue for Hispanics. Nonetheless, the modified two-group model may provide a plausible alternative scenario about where truth lies regarding the total (male and female combined) effect of correlation bias on coverage estimates for Hispanics.

The results for Hispanics in Table 3 show that the correlation bias adjustment in A.C.E. Revision II (second column) has a relatively mild effect (yielding an estimated undercount of 0.71 percent rather than the 0.42 percent undercount from estimates without a correlation bias adjustment). In contrast, assuming that Hispanics have the same correlation bias as Blacks has a dramatic effect, increasing the estimated undercount for Hispanics to 3.17 percent. This result raised concerns that the much lower undercount estimate for Hispanics from A.C.E. Revision II may reflect error in the estimate of correlation bias because of the limitations of the data and the methodology used in obtaining that estimate.

Synthetic Estimation Error and Choice of Post-Strata

The population of the United States is comprised of many communities living in many different areas and jurisdictions. To some extent each is different. However, in order to calculate coverage estimates, the Census Bureau groups people into a relatively small number of estimation cells, known as post-strata. In forming these estimation groups, the Census Bureau took into account many factors, including race, Hispanic origin, tenure (whether the housing unit was owned or rented), and census mail return rates, as well as age and sex. Research has shown that these variables, known as post-stratification factors, help explain variations in the coverage of the population. (Although factors related to geography can be used, limitations on the sample size of the coverage survey preclude defining detailed geography as part of the post-stratification.)

This method of calculating coverage estimates assumes that all people within each of these estimation cells have the same coverage rate. As a result, the actual estimated undercount rate for a particular area is determined by the characteristics of the people who live in that area and their estimated coverage rates. The final results, known as synthetic estimates, are estimates of the population in small areas that are corrected for census coverage errors.

Clearly, all people within each of these estimation groups have, at best, only approximately the same census coverage rate. Since synthetic estimation is always a compromise between the nearly infinite variety of the actual population and the finite number of possible estimation cells, synthetic estimates will always be, to some extent, inaccurate. At worst, the true coverage may vary greatly within these estimation cells and may depend heavily, in some cases, on local conditions. When the stratification factors fail to capture real and significant variations between local areas, the synthetic estimation method will not work well. In general, the smaller the geographic area, the more likely it is that unusual local conditions will lead to larger amounts of synthetic estimation error. In particular, errors from synthetic estimation are likely to have a larger impact on the estimates for areas with population under 25,000, which include 29 percent of the U.S. population of incorporated places, and constitute 94 percent of the places for which intercensal population estimates are produced.

To refine this methodology for A.C. E. Revision II, the Census Bureau developed one set of estimation cells (or post-strata) to estimate census omissions and a somewhat different set to estimate census

erroneous enumerations. Characteristics used in both sets of estimation cells included age, sex, race, ethnicity, and tenure. In addition, the set used to estimate omissions included characteristics such as metropolitan status and type of enumeration area, mail return rates, and region. The set used to estimate erroneous enumerations included proxy status in the census, household relationship and size, and type and date of return.

This approach of using two sets of estimation cells was tried for the first time in the A.C.E. Revision II estimates in an attempt to reduce synthetic error. It was realized in hindsight that a technical problem with the approach could actually have led to systematic biases in the estimates and, consequently, an increase—not a reduction—in errors from synthetic estimation. The technical problem resulted not fundamentally from the use of the two sets of post-strata, but rather from the fact that some of the new factors used in the post-stratification for estimating erroneous enumerations (particularly proxy status in the census, the most effective of these factors) could not readily be used (or even tested) for the post-stratification for estimating census omissions. Without going into detail (for which, see the Technical Assessment of the A.C.E. Revision II Estimates), whether (or not) including proxy status in the post-stratification increased (or decreased) synthetic error depended on how the omission rates for proxies (the probabilities those persons actually had of being missed by the census) compared to those for non-proxy responses. Since this comparison is unknown, whether using separate estimation cells increased or decreased synthetic estimation error in the A.C.E. Revision II estimates is not known. What is known is that the post-stratification produced estimates of relatively large overcounts for some small places and a few small counties (particularly those with high levels of proxy response to the census), and that the validity of the more extreme estimates is in question.

To summarize, two concerns regarding synthetic error in the A.C.E. Revision II estimates are important. One is the general concern about the level of error in the synthetic estimates because they fail to account for unusual local variations in census coverage. The second is the technical problem with the use of factors to define the estimation cells for erroneous enumerations that could not be readily used (or tested) in defining the estimation cells for omissions. The latter concern leaves uncertainty about whether the separate post-stratification actually decreased or increased error from synthetic estimation, and raises troubling questions about the validity of the more extreme overcount estimates.

Comparison with Demographic Analysis

An important component of reviewing the A.C.E. Revision II coverage estimates is comparing them with the corresponding estimates based on demographic analysis. The A.C.E. Revision II estimate of a total population of 280.1 million is 1.7 million below the demographic analysis estimate. With a Census

2000 count of 281.4 million, A.C.E. Revision II implies a net overcount of 1.3 million, or 0.48 percent, compared with a net undercount of 0.12 percent using demographic analysis.¹

In part because of the correction for correlation bias of Black men, net undercount rates by sex and Black/non-Black categories are now, in aggregate, roughly similar between demographic analysis and A.C.E. Revision II results. Net undercount rates for Black men are 4.19 percent in A.C.E. Revision II and 5.15 percent in demographic analysis. For Black women, A.C.E. Revision II estimates a 0.61 percent net overcount, while demographic analysis shows a 0.52 percent net undercount. Non-Black men have a net overcount of 0.19 percent in A.C.E. Revision II and a net undercount of 0.21 percent in demographic analysis. Results for non-Black women are similar, with a net overcount of 1.41 percent in A.C.E. Revision II and a net overcount of 0.78 percent in demographic analysis.

However, the A.C.E. Revision II estimates and the demographic analysis results are not consistent regarding coverage rates for children aged 0 to 9. While demographic analysis shows a relatively large net undercount of 2.56 percent for children aged 0 to 9, the A.C.E. Revision II estimate is not significantly different from zero. The demographic analysis estimate for children aged 0 to 9 is based largely on birth statistics, which are of high quality in recent years. Thus, these inconsistent findings are particularly puzzling and may indicate an undiscovered problem in the A.C.E. Revision II estimates. One possible explanation may be the lack of correction for correlation bias for children.

Total Error Model and Loss Function Analysis

The Census Bureau has previously used loss function analysis to compare the relative accuracy of the census and the coverage-adjusted estimates. The loss functions use results of a “total” error model that attempts to account for systematic biases that might have been omitted from the coverage correction estimates, as well as variances due to sampling and other random errors. None of the biases arising from the limitations discussed above could be estimated, however, so they could not be incorporated into the current loss function analysis. The omitted biases include errors in the model used to correct for correlation bias, errors from synthetic estimation, and any errors reflected in the inconsistency between demographic analysis and the A.C.E. Revision II results for children aged 0-9. Consequently, the loss function results have not been relied on in deciding about the appropriate population base to be used for intercensal population estimates.

CONCLUSIONS

¹This estimated net overcount from A.C.E. Revision II of 0.48 percent differs from the net overcount estimate of 0.49 percent cited earlier, because it is for the entire resident population, including people in group quarters, in order to be comparable to this demographic analysis result.

The concerns raised above about some of the technical limitations of the A.C.E. Revision II estimates led to the Census Bureau's decision not to adjust the population base for the intercensal estimates. However, the results of A.C.E. Revision II will be used to inform ongoing research on improving the intercensal population estimates. The insights gained on residence issues, the measures of differential coverage, and the implications for measuring immigration will be invaluable in addressing the integration of census, survey, and administrative data that are an integral part of the intercensal estimates program.

Table 1. Distribution of Incorporated Places by Population Size, 2000

Population Size	Number of Places	Percent of Total	Cumulative Percent
Total	19,269	100.0	100.0
>100,000	211	1.1	1.1
25,000 to 99,999	888	4.6	5.7
10,000 to 24,999	1,266	6.6	12.3
2,500 to 9,999	3,316	17.2	29.5
1,000 to 2,499	3,167	16.4	45.9
259 to 999	5,423	28.1	74.1
100 to 249	2,727	14.2	88.2
<100	2,271	11.8	100.0

Source: U.S. Census Bureau, Technical Assessment of A.C. E. Revision II, Table 10.

Table 2. Distribution of Incorporated Places by Population Size and Estimated Net Census 2000 Coverage

	Population Size									
	All Sizes		>100,000		10,000-99,999		1,000-9,999		<1,000	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All Places	19,269	100.0	211	100.0	2,154	100.0	6,483	100.0	10,421	100.0
Places with net overcount										
Subtotal	15,402	79.9	85	40.3	1,544	71.7	4,964	76.6	8,809	84.5
2 percent or more	6,733	34.9	0	0.0	126	5.8	1,268	19.6	5,339	51.2
0 to 2 percent	8,669	45.0	85	40.3	1,418	65.8	3,696	57.0	3,470	33.3
Places with net undercount										
Subtotal	3,867	20.1	126	59.7	610	28.3	1,519	23.4	1,612	15.5
0 to 2 percent	3,450	17.9	124	58.8	586	27.2	1,370	21.1	1,370	13.1
2 percent or more	417	2.2	2	0.9	24	1.1	149	2.3	242	2.3

Source: U.S. Census Bureau, Technical Assessment of A.C.E. Revision II, Table 10.

Table 3: Percent Net Undercount Rates for Major Groups, by Model Used to Correct for Correlation Bias

	No Correlation Bias Adjustment	A.C.E. Revision II (Two-Group)	Fixed Relative Risk Model	Prithwis Model	Modified Two-Group Model
Characteristic	Estimate (%) (S.E.)	Estimate (%) (S.E.)	Estimate (%) (S.E.)	Estimate (%) (S.E.)	Estimate (%) (S.E.)
	<i>Column 1</i>	<i>Column 2</i>	<i>Column 3</i>	<i>Column 4</i>	<i>Column 5</i>
Total	-1.12 (0.20)	-0.49 (0.20)	-0.49 (0.20)	-0.49 (0.20)	-0.36 (0.20)
Race/Origin Domain					
Non-Hispanic White	-1.53 (0.20)	-1.13 (0.20)	-1.17 (0.20)	-1.10 (0.20)	-1.39 (0.20)
Non-Hispanic Black	-0.53 (0.41)	1.84 (0.43)	1.84 (0.43)	1.84 (0.43)	1.84 (0.43)
Hispanic	0.42 (0.44)	0.71 (0.44)	0.89 (0.44)	0.58 (0.43)	3.17 (0.49)
Non-Hispanic Asian	-1.12 (0.68)	-0.75 (0.68)	-0.64 (0.70)	-0.72 (0.68)	-1.01 (0.68)
Native Hawaiian and Other Pacific Islander	1.81 (2.73)	2.12 (2.73)	2.47 (2.90)	0.53 (2.26)	1.90 (2.73)
American Indian or Alaska Native on Reservation	-1.16 (1.53)	-0.88 (1.53)	-0.63 (1.57)	-0.97 (1.52)	-1.08 (1.53)
American Indian or Alaska Native off Reservation	0.30 (1.35)	0.62 (1.35)	0.71 (1.38)	0.64 (1.37)	0.39 (1.35)

Source: Appendix A, Table 1, Shores, Roger (2002), "A.C.E. Revision II: Adjustment for Correlation Bias," DSSD A.C.E. Revision II Memorandum Series PP-53, U.S. Bureau of the Census, December 31, 2002.

Note: All net undercounts are for the household population. A negative net undercount denotes a net overcount.